

## **LEVELS OF SERVICE**

### **BACKBONE DATA CENTER PERSPECTIVE**

#### Questions for Service Providers

1. Is our approach (minimum, recommended, desired) to setting levels of service appropriate from your perspective? Are there better approaches you could suggest?
2. Are the minimum levels of service for this provider type adequate to meet your research needs?
3. Are there changes you would suggest in the classification of this provider type's levels of service as minimum, recommended, desirable?
4. Are there additional or better levels of service for this provider type, or any that seem unreasonable to you as a user?
5. Is the mapping of requirements and levels of service to this provider type reasonable? Does the mapping reflect accurately the role of this provider type?
6. How are these services applied?
7. Are these services discipline specific?
8. What guidelines would you like to see written into future AOs (for any type of service provider) to ensure you (user/producer) obtain the things you need to perform your research, applications, or missions?

## 1. Role of the Backbone Data Centers

The group discussed the role of the Backbone Data Centers. The group recognized that the functionality could be characterized by varying degrees of accountability depending on the nature of the data to be managed. The accountability increases as the level of data approaches the source data.

The group recognized that in the present Formulation Team documentation, the Mission Data Center, Multi-Mission Data Center, and Science Data Center activities will provide data to the Backbone Data Centers for archive and distribution to the general public, often after the end of the generating data center project.

The group recognized that the Backbone Data Centers will likely have responsibility for Long-Term Archiving of NASA ESE data in the near term, until the Long-Term Archive is available for the ingest of the NASA data.

The group recognized that the Backbone Data Centers would have principal responsibility for the stewardship and preservation of the Level 0 – Level 1 data from the ESE Program, and for the processing of Level 0 data to Level 1. Additionally, the Backbone Data Centers may provide a failsafe capability for mission essential data products that may not be provided by the designated source.

The group recognized that the Level 1 data contains the engineering investment of the community, while the Level 2 data contains the scientific investment of the community.

The group discussed the relationship between the hierarchy of functionality and levels of accountability and concluded that with increasing responsibility for the stewardship of the lowest levels of data, data centers must be held to higher levels of accountability. The issue of sustainability of data center activities is of concern in the light of the stewardship over time.

## 2. Data Levels

The group observed that data levels are often blurred and depend on the data set. Level 1, Level 2 may not be the appropriate distinction when relating levels of service to data products.

The group observed that when multiple levels of service are offered in response to a proposal process, the minimum level would be the only one considered for funding. In general, the higher levels of service have been derived from the expectations of the scientific community, and this community's expectations will remain at the highest levels. The scientific community support for budgets to

permit the continuation of higher level of services is questionable. Thus there is no need for multiple levels. Rather, the group observed that AOs should be structured to address the criticality of data and services, and to allow responses from diverse communities for differing services to be provided. In structuring the AO, NASA must include evaluation of data services as a part of the evaluation of the AO responses. The evaluation of AO responses should include a process to evaluate the scaling of data services in recognition of the diverse community and instrument specific needs. It is recommended that AOs request that services be provided from a “cafeteria” list or catalog of potential services depending on the needs of the particular community involved in the data.

The group recognized the need to scale the services provided by data center activities according to the size and diversity of the user communities. These services must meet both the needs of the scientific community and those of the general public. In regard to education and outreach activities included in the AO process, the data centers need to identify the list of services needed to enable the outreach and educational activities.

The group recognized that services will migrate over time and should evolve with technology. An example of this is the use of the World Wide Web in providing data and information to users. However, services must be robust under changes in technology. Users need to have stability in services to permit an efficient transfer of data and information, and they need to benefit from the efficiencies provided by changing technology.

### 3. Process

The question was raised as to whether there are combinations of functions that should be requested by NASA procurement processes to insure long-term stability to those activities, such as stewardship of data, including Long-Term Archiving. As an example, should the Backbone Data Center infrastructure be procured through a different process than the data centers that may have less data stewardship responsibilities to insure long-term stability of the Backbone functionality. Such an RFP might include requirements that delineate specific services to be provided, and allocate costs for their provision.

If certain services are procured through specific direction of requirements, what are the responsibilities and accountabilities associated with these procurements.

### 4. Functional / Cost Areas

#### a. Ingest

The services associated with the Ingest function at a Backbone Data Centers are:

Capture the ingest data in a timely manner. The timeliness depends on the data set, and the user community. For Numerical weather prediction, the time delay would be approximately 3 hours from data observation. For scientific data, the time delay may range from days to weeks. If the data is flowing from a satellite, the data must be captured in real time without loss.

Capture the data in a reliable manner. For data captured from the satellite, there must be high reliability and no loss of data. For data transferred from a data provider, the reliability will be determined by the ICD between the data provider and the data center providing ingest. Thus there will be a wide range of reliability requirements on the data center, and a consequent accountability for the data.

The data center must account for the data, including the verification that all data available for ingest has in fact been ingested.

Provide QA for all ingested data, including verification of the transmission of the data to the data center, and the reliability of the ingest process.

Provide for the ingest of all data, metadata, documentation and information about the data.

Provide for the ingest of data from diverse sources, depending on the nature of the Data Center, and of the community that they serve.

Provide for the conversion of data formats from the data format provided by the data generation system, and that of the archive system as appropriate, depending on the community and the future use of the data.

#### b. Processing / Reprocessing

The services associated with the Data Processing function at a Backbone Data Centers are:

Integrate and test the data production algorithms in the data production system.

Maintain the data production algorithms, and all production string scripts and ancillary processing software in the current hardware and software environment.

Provide a schedule driven production environment as contrasted to a rate driven processing system.

Provide documentation of the data processing history, including provenance tracking, algorithm version tracking, ancillary data versioning, operational data quality tracking and data accounting.

Provide for the timely production of data according to ICDs between the data users as appropriate. This will be determined by the operational need for data products, and by the needs of users for the generation of higher-level data products.

Provide for the reprocessing of the data as appropriate, and as determined by the user community.

Enable the leveraging of SEEDS assets to provide services to other agencies for the production of data that may be required for other purposes, including on-demand or near-real time processing for assistance in National emergencies.

Provide the ability to respond rapidly to the need for additional processing capability to offset the loss of assets that may not be able to fulfill needed requirements.

Provide for on-demand processing to reduce the demand load on data distribution services from the data center. This may include data mining activities, name-based and value-based subsetting, object identification, etc.

#### c. Documentation

The services associated with the Documentation function at a Backbone Data Centers are:

Provide and maintain FGDC-compliant metadata for all data sets according to community approved data models as the interface between the scientific community and the data.

Provide and maintain in the current technological environment, read software that will allow the user to readily access and use the data.

Provide and maintain in the current technological environment data documentation, including catalog interfaces and users guides. Provide and maintain in the current technological environment all mission documentation relevant to the data, instruments, and production

environment and history, including the Algorithm Theoretical Basis Document, Data Software Interface Specifications, and other relevant documentation.

Provide and maintain in the current technological environment tools for data access as appropriate.

Maintain the documentation through periodic refresh and update of the documentation technology on a 4 – 5 year basis as determined by best practices in the industry.

#### d. Archive

The services associated with the Archive function at a Backbone Data Centers are:

Provide archive a management environment that is consistent with current industry practice. The NARA standards are recognized as being too rigid and inflexible for an archive that must keep pace with the scientific community in both its practices and its technology.

Provide data backup and a deep archive capability depending on the tradeoff between reprocessing as required to maintain the archive and the cost of maintenance of the deep archive facility.

#### e. Data Distribution

The group recognized that the Backbone Data Centers will have responsibility for the archiving and distribution of the data from the Mission Data Centers, Multi-Mission Data Centers and Science Data Centers after these data centers have concluded their data archive and distribution functions, and for distribution to the general public.

The services associated with the Data Distribution function at a Backbone Data Centers are:

Provide for synchronous on-line data transfer to permit the use of electronic delivery of subsetting data products to the user community. The provision of asynchronous electronic delivery precludes the use of current on-line access software technologies.

#### f. User Services

The services associated with the User Services function at a Backbone Data Centers are:

Outreach is defined as working with potential new users.

User Support is defined as responding to existing users.